MODUL 9

KLASIFIKASI : DECISION TREE

9.1. Tujuan

1. Mahasiswa mampu menggunakan dan membuat model klasifikasi dengan teorema pohon keputusan.
2. Mahasiswa mampu menerapkan algoritma pohon keputusan terhadap studi kasus tertentu.

9.2. Landasan Teori

Decision tree merupakan salah satu metode belajar yang sangat populer dan banyak digunakan secara praktis. Metode ini merupakan metode yang berusaha menemukan fungsi–fungsi pendekatan yang bernilai diskrit dan tahan terhadap data-data yang memiliki kesalahan (*noisy data*) serta mampu mempelajari ekspres-ekspresi *disjunctive* seperti ekspresi OR. *Iterative Dychotomizer version* 3 (ID3) adalah salah satu jenis *decision tree* yang umumnya digunakan untuk menemukan aturan yang diharapkan bisa berlaku umum untuk data-data yang tidak lengkap atau yang belum pernah kita ketahui. Salah satu varian lainnya adalah J48.

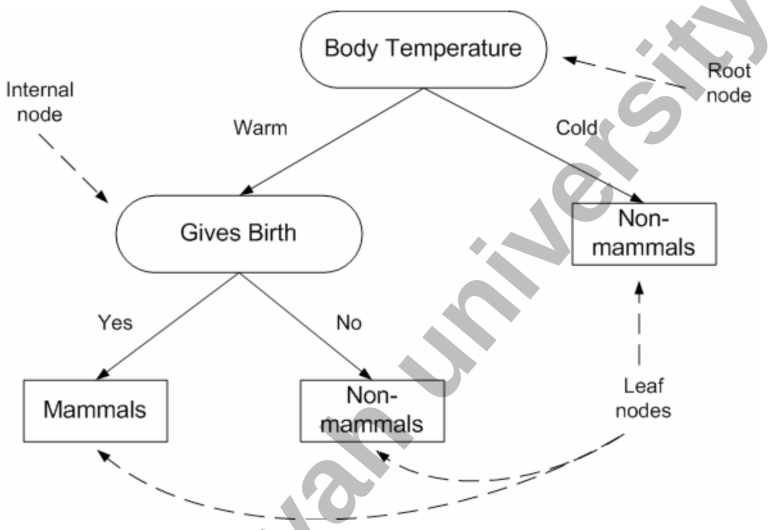
Pohon (*tree*) adalah sebuah struktur data yang terdiri dari simpul (*node*) dan rusuk (*edge*). Simpul pada sebuah pohon terdiri dari 3:

1. Simpul akar (*root node*)
2. Simpul percabangan/internal (*branch/internal node*)
3. Simpul daun (*leaf node*)

©Muhammadiyah university Press

Pohon keputusan merupakan representasi sederhana dari teknik klasifikasi untuk sejumlah kelas berhingga. Simpul internal maupun simpul akar ditandai dengan nama atribut, rusuk diberi label nilai atribut yang mungkin, sedangkan simpul daun ditandai dengan kelas-kelas yang berbeda.

Objek (*record*) diklasifikasikan dengan mengikuti suatu jalur (*path*) yang dimulai dari simpul akar sesuai dengan nilai atribut dalam *record* tersebut.



Gambar 9.1 Contoh Klasifikasi Pohon Keputusan Binatang Mamalia

Pada kegiatan praktikum ini, kita menggunakan pohon keputusan untuk membuat klasifikasi data cuaca terhadap kegiatan bermain tenis. Metode ini tidak akan kita gunakan untuk melakukan prediksi terhadap data uji, namun hanya untuk membuat kelompok-kelompok klasifikasi berdasarkan nilai kelas datanya.

Dengan metode ini, suatu data uji dapat diklasifikasikan berdasarkan kelas data dalam sebuah atribut.

©Muhammadiyah university Press

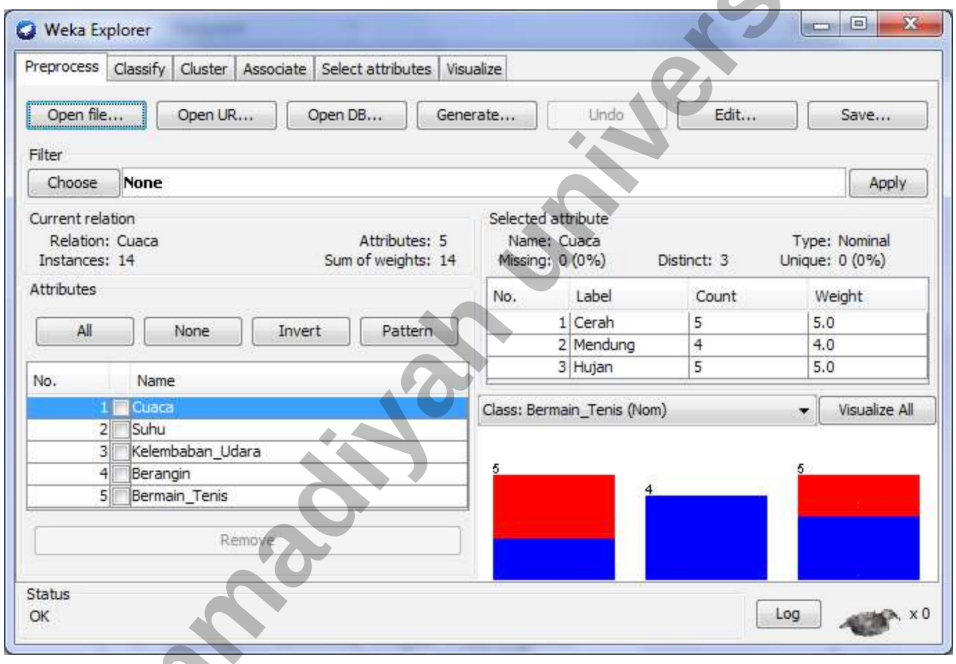
9.3. Alat dan Bahan

1. Komputer dengan sistem operasi Windows.
2. Program aplikasi Weka, RapidMiner.
3. Modul Praktikum Data Warehousing dan Data Mining.

9.4. Langkah-langkah Praktikum

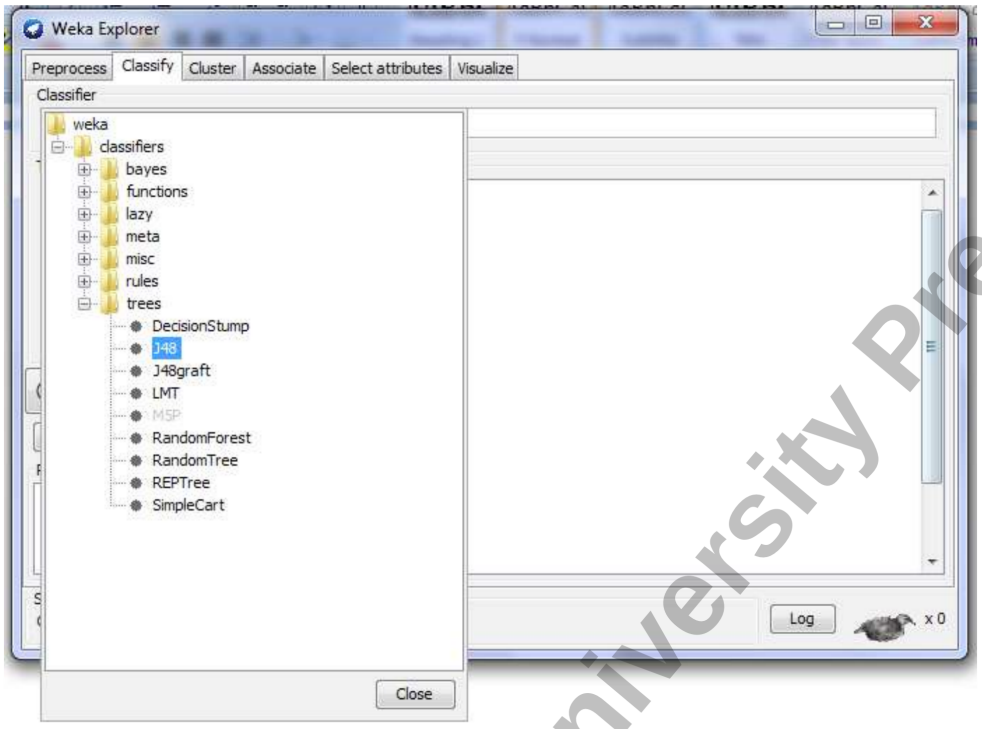
9.4.1. Pohon Keputusan Menggunakan Weka

1. Jalankan aplikasi Weka - Explorer.
2. Buka file **Cuaca.arff**, dengan Weka Explorer.

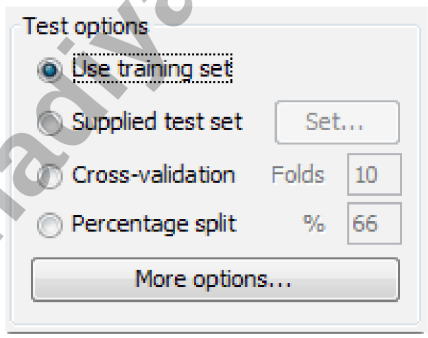


3. Setelah membuka file tersebut, buka tab **Classify**.
4. Tekan tombol **Choose – Trees – J48**. Pada kegiatan ini kita akan menggunakan algoritma pohon keputusan J48.

©Muhammadiyah University Press

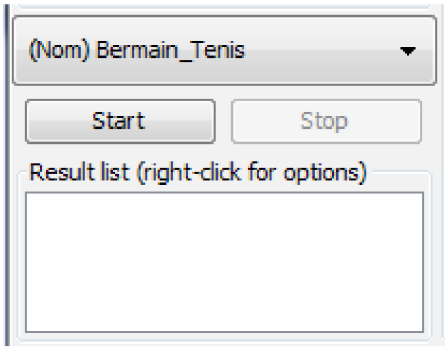


5. Pada pilihan **Test Options**, pilih **Use training set**. Data Cuaca.arff hanya kita gunakan sebagai data training untuk menciptakan klasifikasi.

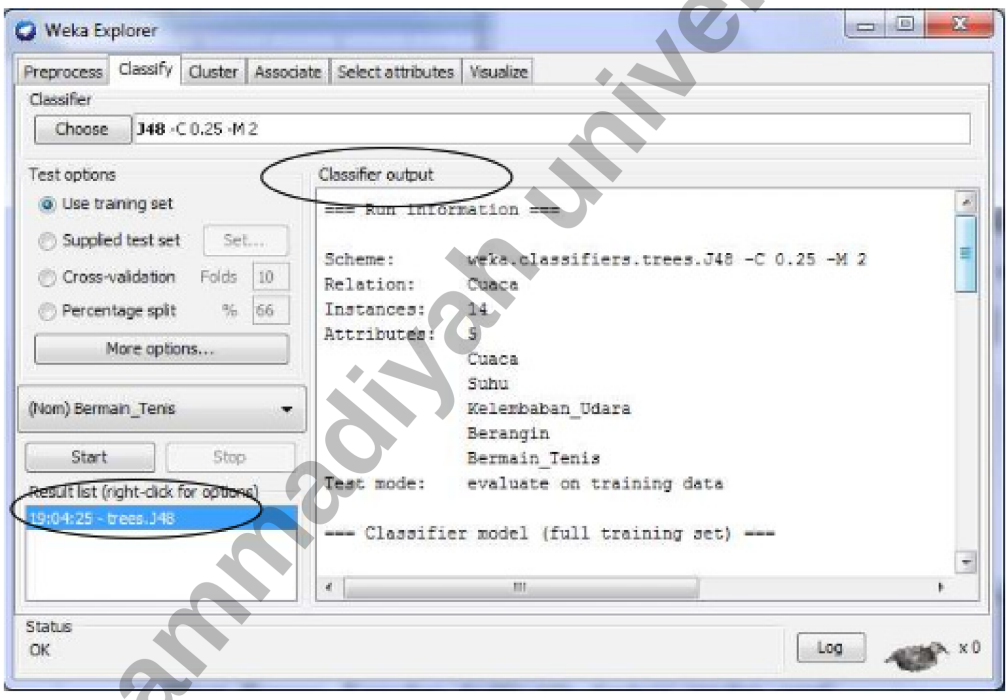


6. Pastikan pada pilihan atribut dependen adalah **Bermain_Tenis**. Kemudian klik **Start** untuk memulai proses klasifikasi pohon keputusan.

©Muhammadiyah University Press



7. Setelah proses selesai, maka akan ditampilkan hasil perhitungan klasifikasi menggunakan algoritma J48. Ada 2 hasil yang diberikan, yaitu pada kolom **Result list**, dan kolom **Classifier output**.



8. Amatilah hasil pada kolom Classifier Output, scroll ke bawah untuk melihat hasil J48 Prunned Tree, waktu yang dibutuhkan untuk proses Training, dan hasil perhitungan klasifikasi.

©Muhammadiyah university Press

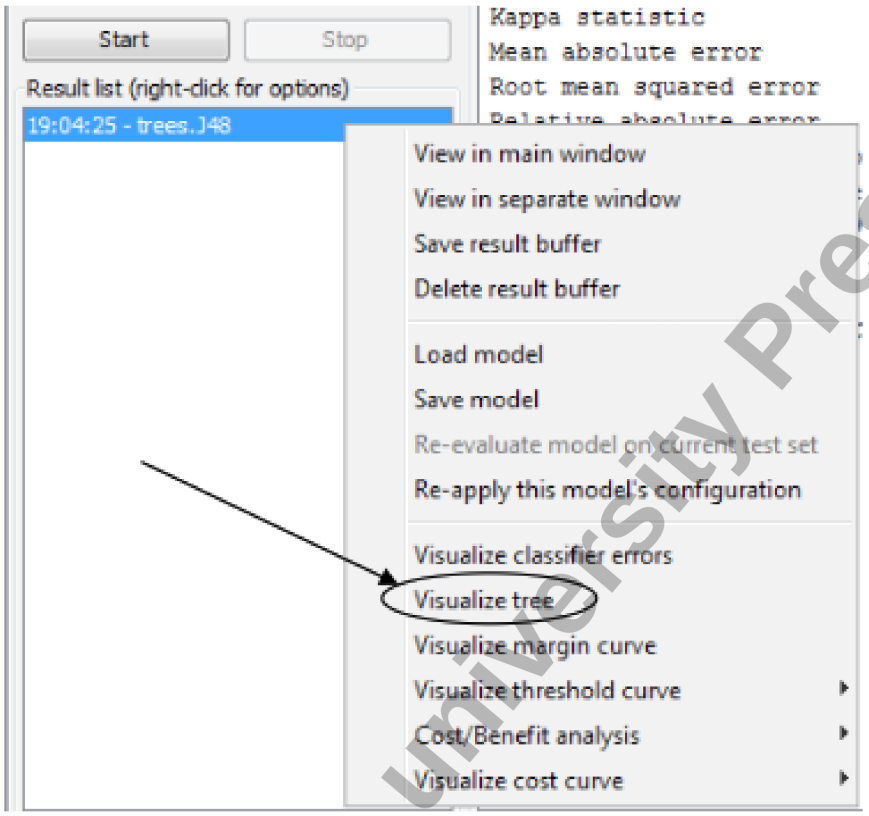
```
J48 pruned tree
------------------

Cuaca = Cerah
|   Kelembaban_Udara <= 75: YA (2.0)
|   Kelembaban_Udara > 75: TIDAK (3.0)
Cuaca = Mendung: YA (4.0)
Cuaca = Hujan
|   Berangin = YA: TIDAK (2.0)
|   Berangin = TIDAK: YA (3.0)

Number of Leaves  :     5

Size of the tree :      8

=== Evaluation on training set ===

Time taken to test model on training data: 0.02 seconds

=== Summary ===

Correctly Classified Instances          14              100%
Incorrectly Classified Instances         0                0%
Kappa statistic                          1
Mean absolute error                      0
Root mean squared error                  0
Relative absolute error                  0       %
Root relative squared error              0       %
Coverage of cases (0.95 level)         100       %
Mean rel. region size (0.95 level)      50       %
Total Number of Instances               14
```

Berdasarkan hasil tersebut dapat diketahui sebagai berikut:

a) Jumlah simpul daun pada pohon keputusan = 5
b) Jumlah simpul keseluruhan pada pohon keputusan = 8
c) Waktu yang dibutuhkan untuk proses pelatihan = 0,02 detik
d) Tingkat ketepatan klasifikasi = 100%
e) Tingkat ketidaktepatan klasifikasi = 0%

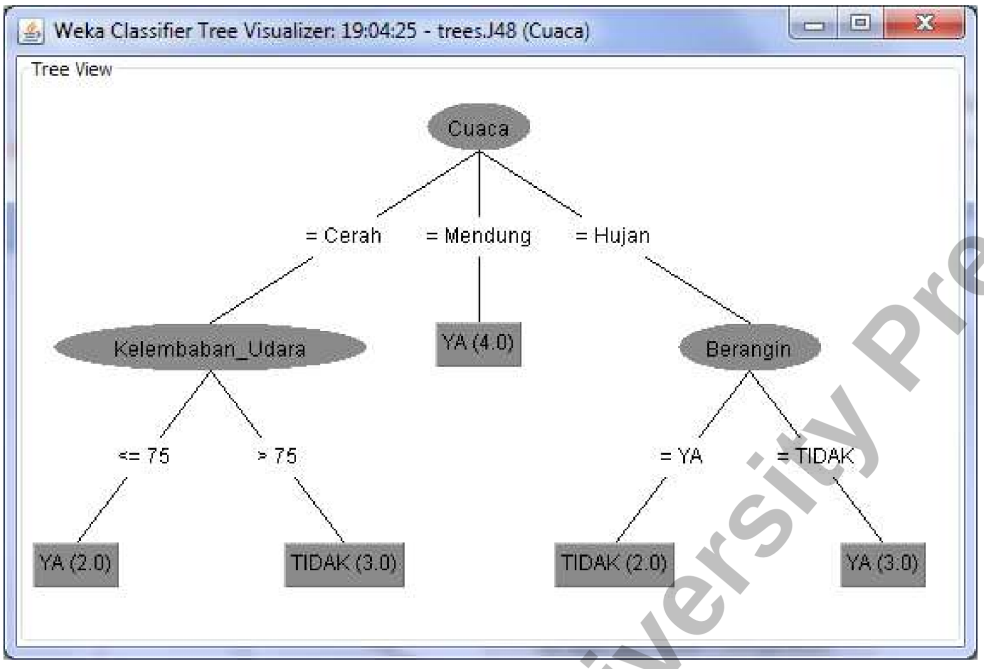
9. Untuk melihat hasil skema pohon keputusan, kembali ke kolom **Result list**. Klik kanan pada hasil **trees.J48 – Visualize tree**.

©Muhammadiyah university Press



10. Sebuah jendela Weka Classifier Tree Visualizer akan ditampilkan. Pada jendela ini akan tampak Tree View, hasil klasifikasi pohon keputusan.

©Muhammadiyah University Press

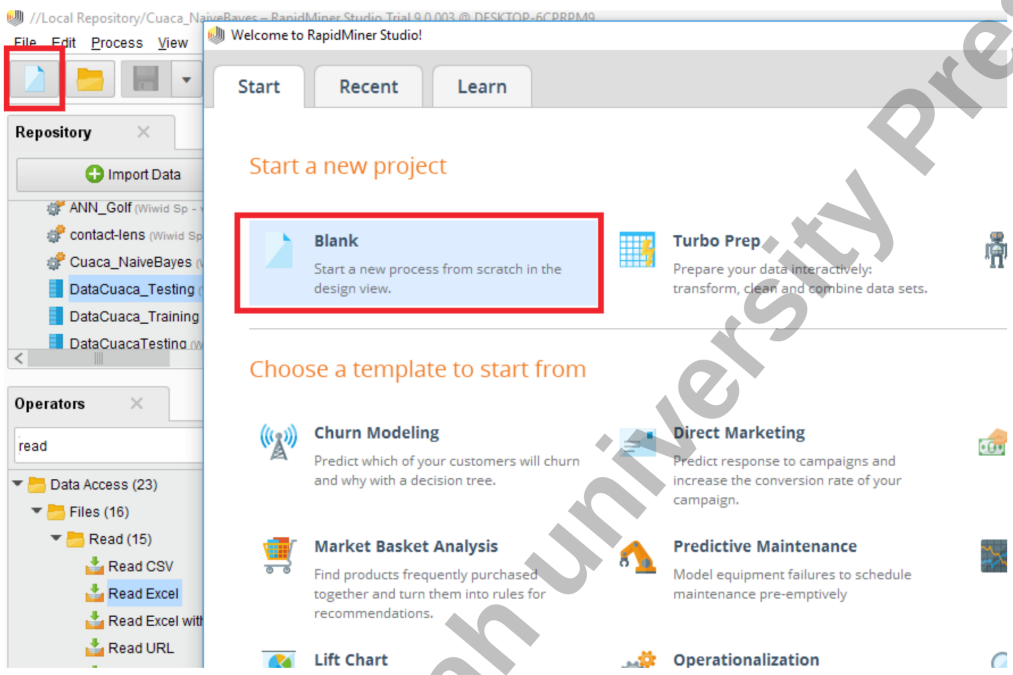


11. Berdasarkan pohon keputusan tersebut, dapat dilihat jenis-jenis simpul yang ada, sebagai berikut:
    a) Simpul akar = Cuaca
    b) Simpul internal = Kelembaban_udara, dan Berangin
    c) Simpul daun = YA, TIDAK
12. Klasifikasi yang terbentuk yaitu:
    a) Seseorang akan bermain tenis (YA) jika kondisi sebagai berikut:
        i. Cuaca = Cerah, Kelembaban_udara <= 75, (nilai atribut lain diabaikan).
        ii. Cuaca = Mendung, (kondisi lain diabaikan).
        iii. Cuaca = Hujan, Berangin = TIDAK, (nilai atribut lain diabaikan).
    b) Seseorang tidak akan bermain tenis (TIDAK) jika kondisi sebagai berikut:
        i. Cuaca = Cerah, Kelembaban_udara > 75, (nilai atribut lain diabaikan).
        ii. Cuaca = Hujan, Berangin = YA, (nilai atribut lain diabaikan).

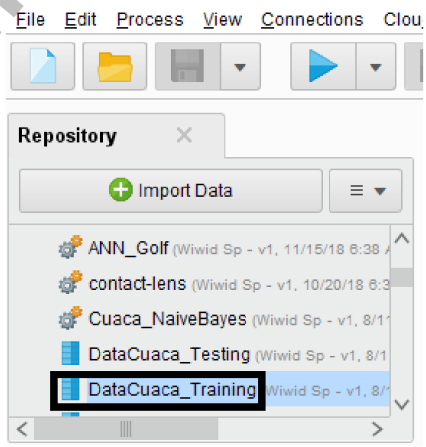
©Muhammadiyah university Press

9.4.2. Pohon Keputusan Menggunakan RapidMiner

1. Buka aplikasi RapidMiner.
2. Klik **New Process** (Ctrl + N) → **Blank** pada halaman perspective RapidMiner.

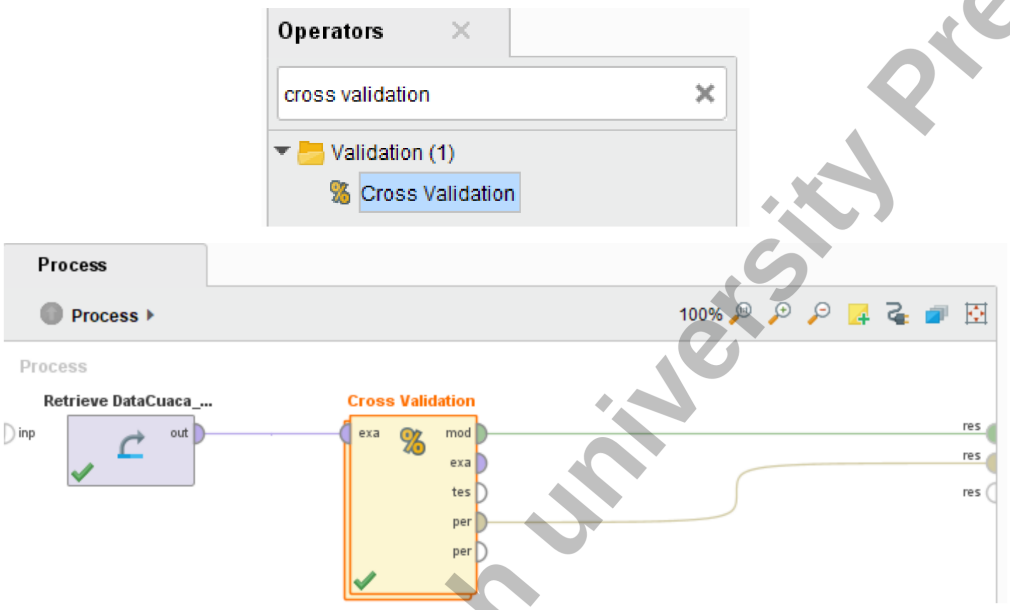


3. Pada area Process View, kita akan mendesain model proses pohon keputusan menggunakan data DataCuaca_Training yang telah dibuat pada kegiatan 8.4.2 dalam Modul 8.

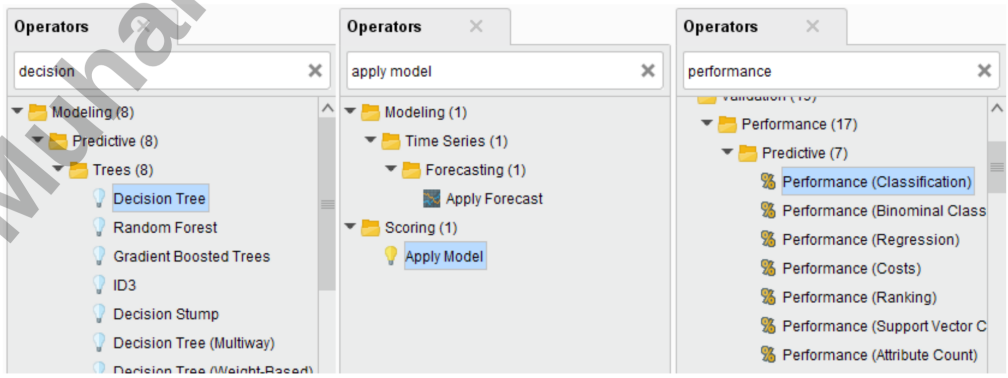


©Muhammadiyah university Press

4. Drag **DataCuaca_Training** ke area Process View. Drag pula operator **Cross Validation** ke area Process View. Hubungkan port output data training ke port input Cross Validation serta port output Cross Validation dengan port input Result seperti gambar berikut.

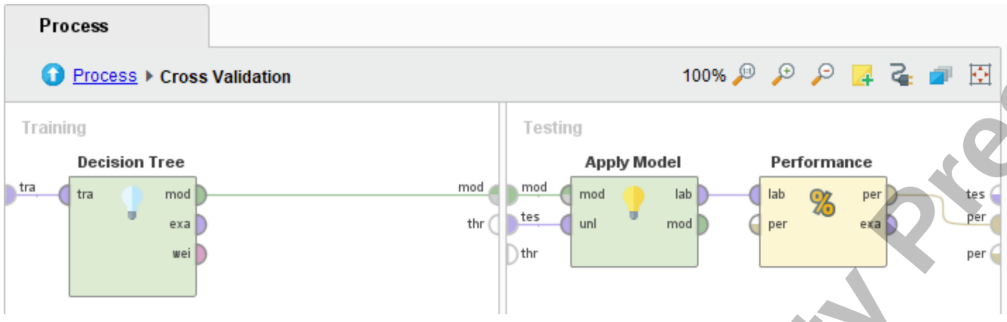


5. Selanjutnya adalah merancang algoritma pohon keputusan. Klik ganda operator **Cross Validation** yang terdapat pada Process View, sehingga masuk ke jendela Process – Validation.
6. Masukkan operator **Decision Tree** dalam area Training, operator **Apply Model** dan **Performance** dalam area Testing.

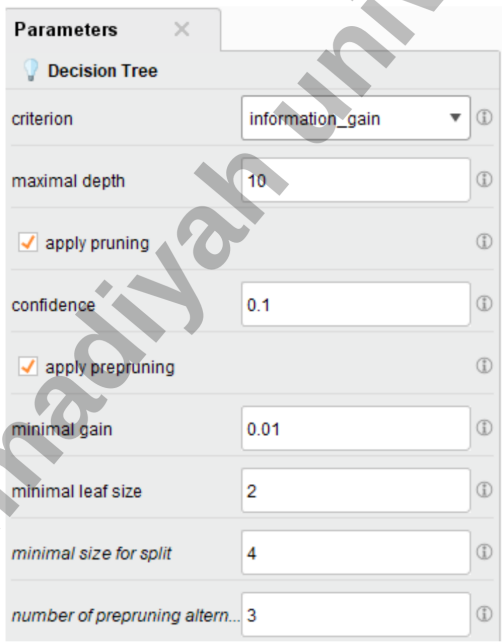


©Muhammadiyah university Press

7. Hubungkan port input dan output masing-masing operator seperti gambar berikut.

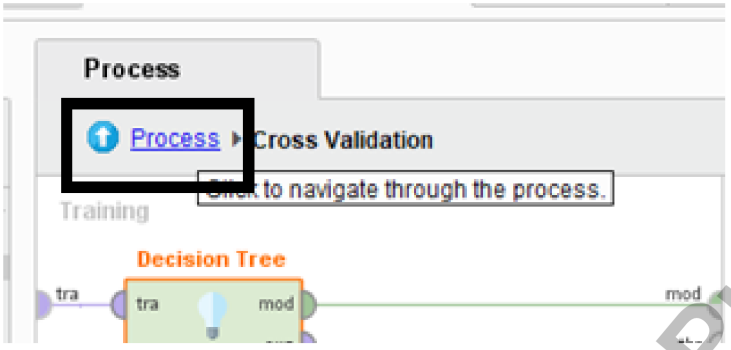


8. Klik Decision Tree pada area Training, dan pastikan kriteria (Criterion) yang dipakai adalah Information Gain (J48) pada kolom Parameter di sebelah kanan Process View.

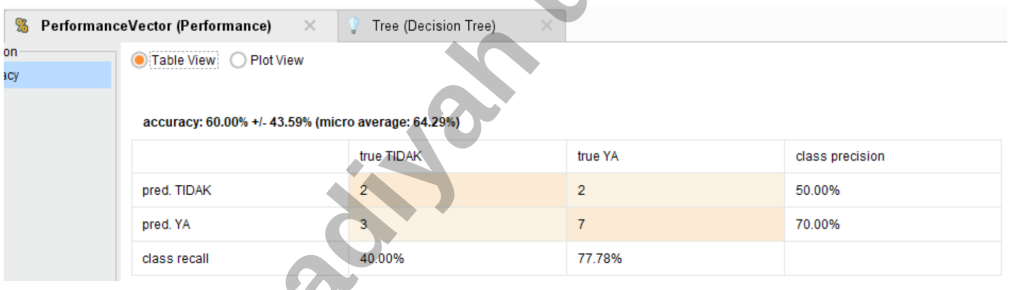


9. Tekan tombol panah ke atas atau tekan tombol Process yang terletak di atas area view untuk kembali ke desain awal.

©Muhammadiyah University Press



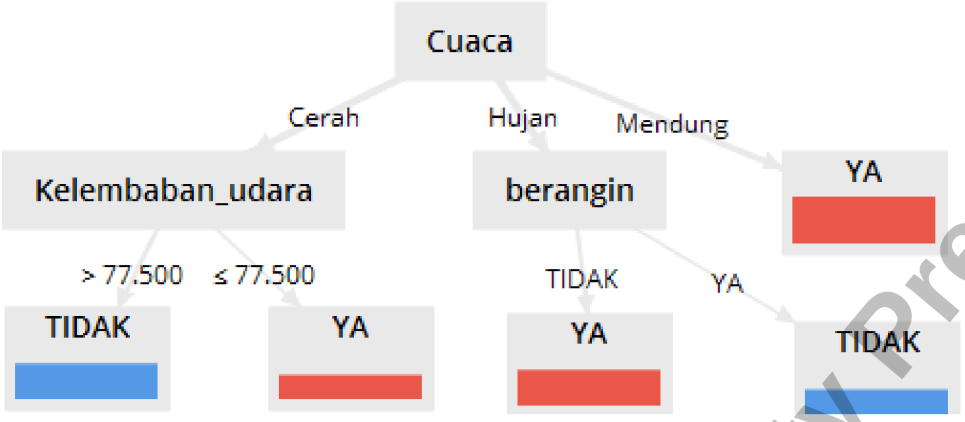
10. Jalankan proses dengan menekan tombol **Run** (atau menekan tombol F11).
11. Perhatikan hasil proses klasifikasi pohon keputusan. Berikut 2 hasil proses klasifikasi pohon keputusan:
    a) PerformanceVector (Performance), menunjukkan tingkat akurasi, presisi, recall dan lain-lain dalam bentuk tabel dan plot view. Bisa anda coba satu persatu untuk melihat.



	true TIDAK	true YA	class precision
pred. TIDAK	2	2	50.00%
pred. YA	3	7	70.00%
class recall	40.00%	77.78%	

    b) Tree (Decision Tree), menunjukkan hasil pohon keputusan dari proses klasifikasi. Dapat dilihat bahwa pohon keputusan yang dihasilkan sama seperti saat menggunakan Weka.

©Muhammadiyah university Press



12. Berdasarkan pohon keputusan tersebut, dapat dilihat jenis-jenis simpul yang ada, sebagai berikut:
    a) Simpul akar = Cuaca
    b) Simpul internal = Kelembaban_udara, dan Berangin
    c) Simpul daun = YA, TIDAK
13. Klasifikasi yang terbentuk yaitu:
    a) Seseorang akan bermain tenis (YA) jika kondisi sebagai berikut:
        i. Cuaca = Cerah, Kelembaban_udara <= 77,5. (nilai atribut lain diabaikan).
        ii. Cuaca = Mendung. (nilai atribut lain diabaikan).
        iii. Cuaca = Hujan, Berangin = TIDAK. (nilai atribut lain diabaikan).
    b) Seseorang tidak akan bermain tenis (TIDAK) jika kondisi sebagai berikut:
        i. Cuaca = Cerah, Kelembaban_udara > 77,5. (nilai atribut lain diabaikan).
        ii. Cuaca = Hujan, Berangin = YA. (nilai atribut lain diabaikan).

©Muhammadiyah University Press

9.5. Tugas

1. Berdasarkan pohon keputusan pada kegiatan 9.4.2 (menggunakan RapidMiner), isikan nilai kelas atribut Bermain_Tenis pada tabel Testing berikut:

Cuaca	Suhu	Kelembaban _udara	Berangin	Bermain_Tenis
Cerah	75	65	TIDAK	.....
Cerah	80	68	YA	.....
Cerah	83	87	YA	.....
Mendung	70	96	TIDAK	.....
Mendung	68	81	TIDAK	.....
Hujan	65	75	TIDAK	.....
Hujan	64	85	YA	.....

2. Gunakan file ARFF yang dikerjakan pada Tugas nomor 1 dalam Modul 7 sebagai data training.
   a) Buatlah dan cetaklah pohon keputusan berdasarkan data tersebut!
   b) Carilah nilai-nilai parameter berikut:
      i. Jumlah simpul daun pada pohon keputusan = ...
      ii. Jumlah simpul keseluruhan pada pohon keputusan = ...
      iii. Waktu yang dibutuhkan untuk proses pelatihan = ... detik
      iv. Tingkat ketepatan klasifikasi = ...%
      v. Tingkat ketidaktepatan klasifikasi = ...%
3. Gunakan file Excel yang dikerjakan pada Tugas nomor 1 dalam Modul 6 sebagai data training.
   a) Buatlah dan cetaklah pohon keputusan berdasarkan data tersebut!
   b) Cetaklah perspektif Plot View dengan model Scatter. X-Axis = Gender, Yaxis = Asisten, dan Color Column = Lama_Studi. Nilai Jitter bisa diubah-ubah untuk memperoleh pola penyebaran yang lebih jelas.

©Muhammadiyah University Press

4. Berdasarkan pohon keputusan dari soal nomor 2, tentukan klasifikasi yang terbentuk berdasarkan kondisinya sesuai dengan simpul-simpulnya.

©Muhammadiyah university Press